**Maximizing Revenue for Taxi Cab Drivers through Payment Type Analysis**

**Problem statement:** to maximise revenue streams for taxi drivers in order to meet this need.

**Objective:** find out if there is a big difference in the fares for those who pay with credit cards versus those who pay with cash.

**Libraries used:** pandas , matplotlib, seaborn, scipy, stats model.

**Columns** : 18

**Rows :** 2,48,575

**Steps:**

1. Data as given by the client through sql database
2. We can Ensure data accessibility and permissions
3. After multiple discussion with stake holders we concluded wih key KPI’s are :

passenger\_count,payment\_type ,fare\_amount,trip\_distance ,duration.  
  
4.we used methodology:

* Descriptive Analysis: Performed statistical analysis to summarize key aspects of the data, focusing on fare amounts and payment types
* Hypothesis Testing : Conducted a T-test to evaluate the relationship between payment type and fare amount

1. Then we did data preprocessing by removing null valuesless than 1% data, outliers treatment with winsorization technique and removing duplicates
2. We visualized using boxplot and scatter plot for correlation coefficient.
3. Then we did eda (explorartory data analysis) process like mean, median , mode , standard deviation, varience , skewnwess and kurtosis
4. then we used auto eda techniques like sweetviz and dtale
5. **In our project**,During the initial analysis, we realized that the existing columns didn't capture all the necessary information. To address this, we engaged in feature engineering to create new columns that would provide more insights.
6. Analyzed taxi pickup patterns to identify highest number of pickups occurred between 5 PM and 7 PM & 6 PM being the peak hour.
7. The results showed a significant preference for credit card payments, which made up approximately 74% of all transactions and cash payments accounted for about 26%. by using pie chart
8. first looked at a histogram and then confirmed it with a QQ plot. This step was important to decide on the right statistical test.
9. **Hypotheses:**

* Null hypothesis: There is no difference in average fare between credit card payments and cash payments.
* Alternative hypothesis: There is a difference in average fare between credit card payments and cash payments.

1. After conducting the test, I found that the p-value was much smaller than the 5% significance level. This means that the difference in average fares between the two payment methods is statistically significant.
2. Customers paying with cards tend to have a slightly higher average trip distance and fare amount compared to those paying with cash.

**Conclusion:**

**Business Insight:**

"This insight is valuable because it suggests that encouraging customers to use credit cards can help generate more revenue for taxi cab drivers."

**Why did we used hypothesis**

**We observed that 74% of customers use credit cards and 26% use cash. To scientifically validate the impact of payment methods on fare amounts, we conducted hypothesis testing**